

Interopérabilité

Plan du cours :

1. Introduction
2. Normes, standards, formes
3. Enjeux de l'interopérabilité
 - A. Enjeux techniques
 - B. Enjeux économiques
 - C. Enjeux de la mondialisation
 - D. Enjeux méthodologiques
 - E. Enjeux pour les usagers
4. Modèles d'interopérabilité
5. Syntaxe et sémantique
6. Interopérabilité et SI en santé
7. Les terminologies

1. Introduction

L'**interopérabilité** ou **interfonctionnement** en informatique = la capacité que possède un système informatique à **fonctionner avec d'autres produits ou systèmes informatiques**, existants ou futurs, **sans restriction d'accès ou de mise en œuvre**. Les deux termes sont normalisés par la CSA (Conseil Supérieur de l'Audiovisuel) et la Commission électrotechnique internationale.

L'interopérabilité est considérée comme **très importante** voire **critique** dans de **nombreux domaines**, dont l'informatique, le médical au sens large, les activités ferroviaires, l'électrotechnique, l'aérospatiale, le domaine militaire et l'industrie en général.



Développer l'interopérabilité consiste à :

- Garantir que **l'échange d'information** puisse se faire sans trop dépendre des logiciels utilisés.
- **Éviter des restrictions d'accès ou de mise en œuvre**, comme l'impossibilité de lire certains formats de fichier par exemple...

En règle générale, cela consiste à **mettre en place des standards et des normes**, clairement établies et univoques (=uniques, dont toujours le même sens). Ces standards fixent des exigences et des recommandations pour que deux systèmes informatiques puissent fonctionner ensemble sans problème.

Un **frein majeur** à une interopérabilité correcte est l'utilisation dans les matériels et les logiciels de formats dont **seuls leurs concepteurs ont les clefs** (comme les logiciels propriétaires).

2. Normes, standards, formats

Norme (du latin norma, équerre, règle) = désigne un état habituellement répandu ou moyen considéré le plus souvent comme une règle à suivre.

→ Le processus d'élaboration d'une norme est appelé normalisation.

Organismes de normalisation = Organismes dont le rôle est de **valider des normes** que les industries utiliseront comme support pour rendre leurs services et produits interopérables.

Voici les plus importantes :

- ISO : Organisation Internationale de Normalisation
- ANSI : American National Standards Institute
- AFNOR : Association Française de NORmalisation
- CEN : Comité Européen de Normalisation
- ITU : Union Internationale des Télécommunications
- CENELEC : Comité Européen de Normalisation ELECtrotechnique



→ Standards ou normes :

En informatique il ne faut pas faire la confusion entre une **norme** et un **standard**, ce dernier désignant ce qui est **produit habituellement par un producteur et ne dépend que de lui** : par exemple les fichiers .DOC ou .PPT sont devenus des ‘standards’ de fichier “texte” ou “présentation”. On parle de **standard de facto**. Cette confusion vient de l’anglais, qui n’a qu’un seul mot pour désigner les deux concepts (standard signifie aussi norme).

Mémo : (je voyais la norme un peu comme une loi et le standard je prenais l’exemple des câbles apple avec le chargeur. Il n’y a qu’apple qui a ces câbles, ils ne sont pas répandus par conséquent c’est un standard.)

L’interopérabilité nécessite que les informations nécessaires à sa mise en œuvre soient disponibles sous la forme de **standards ouverts**.

Un format ouvert = format de données dont les spécifications techniques sont **publiques** et **sans restriction d’accès** ni de mise en œuvre, par opposition à un format fermé. (*loi française n° 2004-575 du 21 juin 2004*)

La norme, et/ou la recommandation qui l’accompagne, est établie (normalisation) par un **organisme indépendant** qui limite les modifications unilatérales.

On comprend donc qu’il est inopportun de définir une interopérabilité informatique à partir d’un standard non ouvert.

Par exemple, la **norme** peut définir des éléments comme :

- Les formats des données échangées dans le contexte considéré, qui décrivent des séquences d’informations ou de commandes qu’un système doit envoyer, comment ses correspondants doivent y répondre (protocole de communication).
- Les tensions et les courants à utiliser ;
- Les types de câbles à utiliser...



Certaines normes de l'ISO pour les TIC (Technologies de l'Information et de la Communication) :

- Données et records management
- Terminologies
- Échanges d'informations et de données
- Langages de programmation
- Réseaux
- Exigences logicielles, évaluation des procédés
- Sécurité des systèmes d'information
- Archivage électronique
- Information géographique
- Recherche d'information
- Sécurité de l'information

Le **JTC1** est l'**organe de référence** pour la normalisation des Technologies de l'Information au niveau mondial. Il réunit les compétences relatives aux **logiciels** (langage de programmation, codage de l'information...) et celles en matière de **matériels** (microprocesseurs, imprimantes...).

Le total de normes publiées par le JTC1 depuis sa création en 1987 était de 1993 en 2006. Le comité comporte **28 pays membres** et **44 observateurs**.

3. Enjeux de l'interopérabilité

A. Enjeux techniques

L'interopérabilité informatique est un des éléments essentiels pour les **réseaux de télécommunication mondiaux** comme le téléphone et l'Internet.

→ Ces réseaux hétérogènes utilisent/ intègrent des matériels, logiciels et applications variés et d'âge différent (legacy systems) pour assurer le stockage, le traitement et le transport des données de l'information.



B. Enjeux économiques

Les enjeux de l'interopérabilité informatique sont **techniques** et surtout **économiques**.

→ Dans presque tous les domaines d'activité, dans l'industrie et dans les services, en agriculture également, des systèmes informatiques gèrent des données, pilotent des systèmes de contrôle, des systèmes de gestion, et sont interconnectés d'une entreprise à l'autre par des réseaux informatiques (réseaux privés, internet, messageries électroniques...).

C. Enjeux de la mondialisation

Du point de vue de **l'ingénierie des systèmes**, c'est l'interopérabilité informatique qui régit en grande partie l'interopérabilité globale (tant sur le plan des méthodes que des outils mis en œuvre).

D. Enjeux méthodologiques

L'interopérabilité informatique nécessite que les communications obéissent à des **normes**, clairement **établies** et **univoques**.

→ Ces documents techniques définissent souvent des **exigences**, parfois accompagnées de **recommandations** plus ou moins optionnelles.

E. Enjeux pour les usagers

Deux systèmes qui satisfont aux exigences d'une même norme doivent pouvoir **dialoguer** (communiquer, échanger) ensemble sans souci particulier.

→ Ils peuvent ainsi **évoluer librement** (mises à jour fonctionnelles pour suivre les évolutions du domaine/métier concerné) sans risque de casser cette possibilité de communication, **tant qu'ils respectent la norme** définissant et structurant leurs interfaces.



4. Modèles d'interopérabilité

Un service = une **description abstraite** de fonctionnalités à l'aide de primitives (commandes ou évènements) telles que demande de connexion ou réception de données.

Un protocole = un **ensemble de messages** et de **règles d'échanges** réalisant un service.

Une interface = moyen concret d'**utiliser le service**. Dans un programme, c'est typiquement un ensemble de fonctions de bibliothèque ou d'appels systèmes.

On peut voir qu'il existe différents niveaux d'interopérabilité :

- Protocoles de réseaux
- Fichiers
- Appel de procédures distantes
- Accès aux bases de données
- Applications

→ Le modèle OSI : (Open System Interconnexion)

Le modèle d'interconnexion en réseau des systèmes ouverts (modèles OSI) de l'ISO est un **modèle de communication entre ordinateurs** (*pas confondre organisme et modèles*)

Il décrit les fonctionnalités nécessaires à la communication et l'organisation de ces fonctions.

Le modèle comporte **7 couches réparties en 2 groupes**.

Il y a une architecture en groupe avec :

- **4 couches inférieures**, qui sont plutôt orientées **communication** et sont typiquement fournies par un système d'exploitation.
- **3 couches supérieures** sont plutôt orientées **application** et plutôt réalisées par des bibliothèques ou un programme spécifique.

Je me disais que 3 est un plus petit chiffre que 4 donc il est en haut de la pyramide, au-dessus.



Les couches inférieures :

1. **La couche « physique »** est chargée de la transmission effective des signaux entre les interlocuteurs. Son service est typiquement limité à l'émission et la réception d'un bit ou d'un train de bit continu (notamment pour les supports synchrones).
2. **La couche de « liaison des données »** gère les communications entre 2 machines adjacentes, directement reliées entre elles par un support physique.
3. **La couche « réseau »** gère les communications de bout en bout, généralement entre machines : routage et adressage des paquets.
4. **La couche « transport »** gère les communications de bout en bout entre processus (programmes en cours d'exécution).

Les couches supérieures :

5. **La couche « session »** gère la synchronisation des échanges et les "transactions", permet l'ouverture et la fermeture de session.
6. **La couche « présentation »** est chargée du codage des données applicatives, précisément de la conversion entre données manipulées au niveau applicatif et chaînes d'octets effectivement transmises.
7. **La couche « application »** est le point d'accès aux services réseaux, elle n'a pas de service propre spécifique et entrant dans la portée de la norme.

5. Syntaxe et sémantique

A. Interopérabilité syntaxique

CORBA (Common Object Request Broker Architecture) = une architecture logicielle, pour le développement de composants et d'Object Request Broker (ORB).

Ces composants, qui sont assemblées afin de construire des applications complètes, peuvent être **écrits dans des langages de programmation distincts**, être exécutés dans des processus séparés, voire être déployés sur des machines distinctes.



Les **services d'interopérabilité** sont : nommage et annuaire des objets, cycle de vie, notification d'évènements, transaction, relations entre objets, stockage, archivage, sécurité, authentification et administration des objets, gestion des licences et versions, etc.

B. Interopérabilité sémantique

(sémantique = qui concerne le sens, la signification)

Méta données (métadata) :

→ “informations sur les données”, “informations descriptives sur les ressources”.

L'utilisation de métadata descriptives et standardisées **améliore la recherche** des informations pertinentes dans un réseau de ressources.

Ontologie :

→ **Description formelle des concepts**, rôles et relations qui existent pour un agent ou une communauté d'agents.

Les ontologies fournissent une compréhension commune d'un domaine qui peut être communiquée, elles jouent un rôle majeur dans les échanges d'information.

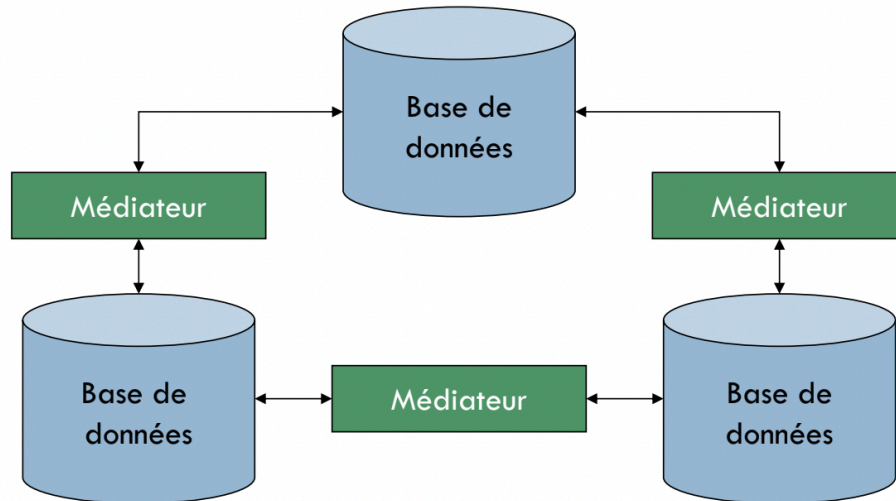
Médiateur :

→ **Adaptateur de données** situé sur un réseau entre un client et un serveur de données (le client peut être une autre base de données).

Un médiateur est un **composant logiciel** qui résout les conflits schématiques et sémantiques.

Exemple de médiateurs : Conversion de supports, d'unités, de structures, Changement de noms, Traduction de noms, Classification d'objets, Regroupements sémantiques...



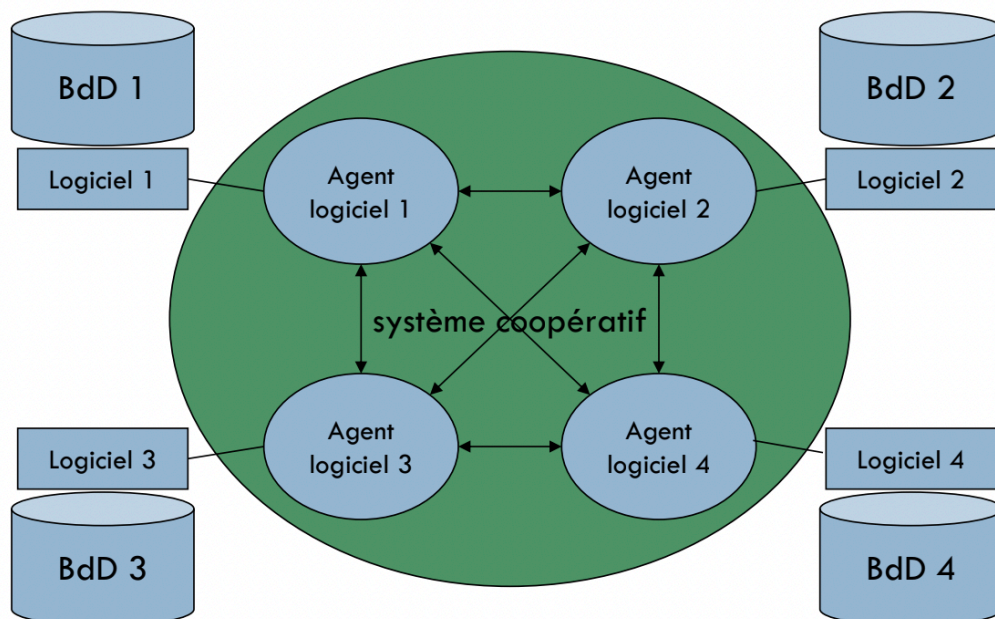


Wrapper :

→ **Composant logiciel** qui fournit des **services d'accès aux données** grâce à un **langage commun** entre les bases de données et les médiateurs ; il assure la traduction des requêtes, met en forme les résultats et les transmet aux médiateurs.

Système multi-agents :

(Un Système Multi-Agents (SMA) comporte plusieurs agents qui interagissent entre eux dans un environnement commun. Certains de ces agents peuvent être des personnes ou leurs représentants (avatars), ou même des machines mécaniques). Ancienne tut



Exemple :

Message M _A	Message M _B
Emetteur : composant A	Emetteur : composant B
Syntaxe B: HL7	Syntaxe B: ASN.1
Cryptage : oui	Cryptage : non
Référentiel : R _A	Référentiel : R _B
Ontologie : O _A	Ontologie : O _B
Nom patronymique : DUPONT	Nom patronymique : DUPONT
Date de naissance : 13/06/1947	Date de naissance : 06/13/47
Urée sanguine: 3,35 mmol/l	Urée sanguine: 10 mg/l
Date de la découverte du cancer : 40 ans	Date de la découverte du cancer : 06/13/87

6. Interopérabilité et SI en santé

HL7 (Health Level 7) = **Standard du domaine "santé"** qui définit un format pour les échanges informatisés de données cliniques, financières et administratives entre systèmes d'information hospitaliers.

→ Initialement américain, ce format s'exporte et tend à s'imposer comme **standard international** pour ce type d'information.

Le chiffre 7 de HL7 indique simplement que c'est un **langage qui fonctionne au niveau de la couche 7**, c'est-à-dire la couche **application**, du modèle OSI. **+++**

→ Cela signifie que HL7 n'a pas à prendre en compte les considérations de sécurité des échanges, ou bien celles du transport du message (ceci étant assuré par des couches de plus bas niveau comme SSL pour la sécurité ou TCP pour le transport des données par exemple).

CI-SIS de l'ANS (Cadre d'Interopérabilité des Systèmes d'Information de Santé (CI-SIS) de l'Agence du Numérique en Santé (ANS)) = **document de référence** qui propose des **règles techniques** et **sémantiques** aux acteurs de santé porteurs de projet devant échanger et partager de la donnée de santé.



Il s'appuie sur des **normes et standards internationaux** et permet de disposer de modèles d'échanges pérennes, tirés des cas d'usage et éprouvés par le processus de concertation. Ce travail évite à chacun de devoir respecifier les données à échanger entre les SI.

Les systèmes d'information dans les domaines sanitaire et médico-social doivent être **communicants** pour **favoriser la coopération des professionnels** dans le cadre des parcours de santé centrés sur le patient (l'utilisateur), et pour aider à la décision médicale.

Le CI-SIS couvre :

- **L'interopérabilité technique**, qui porte sur le **transport des flux** et sur les **services garantissant l'échange et le partage des données de santé** dans le respect des exigences de sécurité et de confidentialité des données personnelles de santé.
- **L'interopérabilité des contenus métiers**, qui permet le **traitement des données de santé et leur compréhension** par les systèmes d'information en s'appuyant sur un **langage commun**.

7. Les terminologies en santé

L'Agence du Numérique en Santé, par l'intermédiaire de son **Centre de Gestion des Terminologies de Santé** (CGTS) est chargée de la **définition**, de la **production**, du **maintien** et de la **publication** de Terminologies sur lesquelles s'appuient les systèmes d'information de santé.

A. Caractéristique des classifications

Le **lien sémantique** qui permet la catégorisation peut être un **lien d'affectation** (type est-un, par exemple, une rubéole est une maladie infectieuse) ou un **lien de partition** (type fait-partie-de, par exemple, le genou fait partie du membre inférieur).

Une classification **monoaxiale** répartit en plusieurs classes disjointes l'ensemble des objets et revient à construire une **hiérarchie de classes** à partir d'une racine unique et commune. Les classes d'un niveau doivent



couvrir l'ensemble du domaine de ce niveau (exhaustivité) sans se recouvrir (exclusivité) afin qu'un objet trouve une place et une seule.

- Très difficile en pratique de répartir les objets de domaine selon un seul critère, ce qui a entraîné le développement de classifications multiaxiales.

Les classifications **multiaxiales** ou à facettes sont modulaires et combinent des termes appartenant à des systèmes différents, eux-mêmes organisés de façon hiérarchique.

- Ainsi, une sémantique de juxtaposition s'ajoute à la sémantique de catégorisation.

Cependant, même la juxtaposition de termes ne suffit pas à exprimer la diversité des liens sémantiques existant entre les termes médicaux (relation de causalité, d'association, d'évolution, etc...)

Aussi ces liens sont-ils explicités par des **connecteurs** (équivalents des groupes verbaux : est un, a, traite, est traité par, cause, est causé par, etc...) dans certaines classifications ou dans les langages artificiels.

En outre, à côté des termes (les mots) et des connecteurs (les verbes), certains systèmes ajoutent des **modificateurs**, faisant fonction d'adjectif, venant **préciser le degré de certitude, l'évolutivité ou l'intensité**.

B. Les classifications

CIM (Classification Internationale des Maladies) :

De l'Organisation Mondiale de la Santé (OMS ou WHO) a été originellement conçue pour coder les causes de décès dans une optique **épidémiologique**. Elle est maintenant également utilisée pour l'évaluation des soins médicaux et l'indexation des dossiers.

Il s'agit d'une **classification monoaxiale** avec 21 chapitres principaux (dont 17 concernent des maladies et 4 concernent les signes et résultats anormaux, les causes de traumatismes, d'empoisonnement ou de morbidité, l'état de santé et les facteurs de recours aux soins).



Les catégories de maladies sont définies en fonction d'un caractère commun qui peut être :

- L'**étiologie** (1= Maladies infectieuses, lettres A et B)
- La **topographie** (9= maladies de l'appareil circulatoire, lettre I)
- La **physiologie** (15= Grossesse et accouchement, lettre O)
- La **pathologie** (II= Tumeurs)

La classification aboutit par subdivisions successives à un code à 3 caractères (une lettre correspondant au chapitre puis 2 chiffres) pour les maladies définies à un niveau général, décliné par l'ajout d'un quatrième chiffre (derrière un point) pour désigner les diagnostics précis et les formes cliniques ; le sous-code 9 désignant l'absence de précision (SAI = sans autre indication) et le sous-code 8 les autres formes non précédemment définies.

Dans certains cas, un cinquième chiffre a été rajouté afin d'améliorer la finesse de description. La CIM-10 a introduit la notion de troubles iatrogènes. Elle compte au total 80 000 termes.

La classification de l'OMS sert en France :

- Au **codage des causes de décès** (codage des diagnostics inscrits sur le certificat de décès)
- Au **regroupement des séjours hospitaliers en groupes homogènes de malades** (programme de médicalisation du système d'information ou PMSI) dont le but est l'analyse médico-économique de l'activité hospitalière.

CCAM (Classification Commune des Actes Médicaux) :

Les classifications des actes et des procédures sont plus variables car elles répondent à différents objectifs.

Le catalogue des actes médicaux par nature (CDAM puis désormais CCAM) a été préparé dans le cadre du PMSI (Programme de Médicalisation des Systèmes d'Information) afin de permettre le codage des actes effectués au cours d'un séjour hospitalier.



SNOMED (=Systematized Nomenclature of Medicine) :

Elle combine une nomenclature de plus de 200.000 termes et une **classification multiaxiale** comportant 11 axes dont : topographie, morphologie, étiologie, altération fonctionnelle, nosologie, actes médicaux.

Chaque axe est défini par une lettre (exemple : T pour topographie, E pour étiologie) et organisé de façon hiérarchique, chaque élément étant associé à un code numérique à 4 ou 5 chiffres.

- ➔ Ainsi un diagnostic est traduit par plus d'un élément signifiant, par exemple la juxtaposition : " T2856 (lobe supérieure du poumon gauche)/ M4100 (inflammation)/ F0300 (fièvre)/ E2012 (pneumocoque) " correspond à la phrase " Pneumonie fébrile à pneumocoque du lobe supérieur gauche ".

L'ajout de connecteurs concernant notamment les liens de causalité permet de décrire un fait complexe en plusieurs phrases.

SNOMED est largement utilisé car précis, cependant ce modèle pose encore des problèmes : les termes des différents axes ne sont pas complètement indépendants entre eux, l'axe Maladie fait souvent double emploi, certains concepts peuvent apparaître dans plusieurs axes.

MeSH (=Medical Subject Headings) :

Elle sert à **indexer, cataloguer** et retrouver des **références de bibliographie** dans le domaine de la Santé. Il a été conçu à la **National Library of Medicine** (NLM) aux Etats-Unis comme support de l'Index Medicus, répertoire des principales publications scientifiques, et est utilisé par les systèmes de recherche bibliographique Medlars et Medline.

MeSH est organisé en **deux parties** : une **liste alphabétique de termes** (lexique) et une **structure multiaxiale**.

➔ Les 200.000 termes du lexique sont distribués selon 15 axes, allant de l'anatomie à la géographie.



→ Les termes équivalents sont rapportés à celui des 20.000 termes principaux (descripteurs) qui exprime le mieux le concept, termes auxquels sont associés un code alphanumérique.

Les **descripteurs** s'organisent selon une **structure hiérarchique**. Des **connecteurs** permettent des références explicites entre termes et expriment les relations de synonymie, de voisinage ou d'association tandis que des **qualificatifs** permettent de considérer les différentes facettes d'un concept (par exemple : tuberculose/traitement).

Ce système indexe actuellement plusieurs centaines de milliers de références et est mis à jour régulièrement pour suivre l'évolution des connaissances.

La multiplicité des classifications répond à la diversité des informations à traduire. Mais parfois, pour représenter la même réalité, les acteurs de la santé utilisent différents systèmes de classification.

Ainsi, pour indexer précisément les dossiers de ces patients en fonction des actes réalisés, un chirurgien utilisera une classification ou un système de mots clés spécifiques de sa spécialité, alors que le codage des actes dans le cadre du PMSI utilise la CCAM.

Cette multiplicité perturbe la vision de l'activité quand elle ne dégrade pas l'information, notamment lors de la communication entre systèmes différents.

Un **transcodage** est une "passerelle" entre deux classifications. La qualité de ces passerelles dépend du type et de la congruence des classifications ainsi reliées. Par exemple, le transcodage de CIM9 en SNOMED ne permet pas toujours une transcription explicite des concepts en raison de la différence de structuration des deux systèmes.

Les transcodages exposent ainsi souvent à une **perte d'information** mais, dans certaines situations, ils sont **impossibles**.

